Modelling perceived risks to personal privacy from location disclosure on online social networks

Fatma S. Alrayes[a], A.I. Abdelmoty[b], W.B. El-Geresy[c] and G. Theodorakopoulos[b]

[a]Information Systems Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; [b]Cardiff School of Computer Science and Informatics, Cardiff University, Cardiff, Wales, UK; [c]Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.

ARTICLE HISTORY
Compiled September 1, 2019

ABSTRACT
As users increasingly rely on online social networks for their communication activities, personal location data processing through such networks poses significant risks to users' privacy. Location tracks can be mined with other shared information to extract rich personal profiles. To protect users' privacy, online social networks face the challenge of ensuring transparent communication to users of how their data are processed, and explicitly obtaining users' informed consent for the use of this data. In this paper, we explore the complex nature of the location disclosure problem and its risks to personal privacy. We evaluate, with an experiment involving 715 participants, the contributing factors to the perception of such risks with scenarios that mimic a) realistic modes of interaction, where users are not fully aware of the extent of their location-related data being processed, and b) with devised scenarios that deliberately inform users of the data they are sharing and its visibility to others. The results are used to represent the users' perception of privacy risks when sharing their location information online and to derive a possible model of privacy risks associated with this sharing behaviour. Such a model can inform the design of privacy-aware online social networks to improve users' trust and to ensure compliance with legal frameworks for personal privacy.

KEYWORDS
Location privacy; Privacy models; Geosocial networks

1. Introduction

As users rely more and more on online social networking applications for their communication activities, the processing of personal location data through such networks increasingly poses significant risks to the security and privacy of users. Such risks stem mainly from the variety of personal identifying data held by these networks and the extended possibility of tracking and profiling users based on their location information. The processing of personal data through such networks is not always transparent to or controllable by the users. On the other hand, the importance of security and privacy of users' data is increasingly being recognised as a challenge to online and mobile applications, as evidenced by the recent personal data leak involving millions of Facebook users (BBC 2018). Also, legal frameworks are emerging that include protection

CONTACT Fatma S. Alrayes. Email: fsalrayes@pnu.edu.sa

mechanisms to allow individuals to better control their personal data. In particular, the General Data Protection Regulation (EU) 670/2016 (GDPR) (EUR-Lex 2018), now in place throughout European Union (EU) countries, stipulates data protection principles and privacy requirements that need to be fulfilled by such applications. To ensure compliance with these legal requirements, privacy awareness methods need to be incorporated into the design of online social networks.

In this work we focus in particular on the processing of personal location data on online social networks. In some types of these networks, denoted *location-based social networks*, users' interaction is mainly guided by their presence in geographic places, e.g. checkins on Foursquare. Processing of location information is essential for the provision of services by these applications. On other networks, denoted *location-enabled social networks*, location is a complementary attribute that can be used to enhance the user experience, e.g. filtering geo-tagged tweets by place on Twitter. Collection and processing of user location information in both cases can result in user profiling and derivation of sensitive information, revealing patterns of presence at home, regularly visited places and frequent activities, and even racial or ethnic origins. For convenience, in the rest of this paper both types of online social networks will be referred to as Geo-Social Networks (GeoSNs).

To comply with legal frameworks and data protection principles, GeoSNs need to observe the *transparency* of user data processing and the *informed consent* of their users for such data processing. In particular, Article 5 of the GDPR stipulates that with respect to transparency, "any processing of personal data should be lawful and fair. It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed ... communication relating to the processing of those personal data be easily accessible and easy to understand ...", whilst Article 7 indicates that processing is based on the freely given and informed consent by the user.

Previous user studies highlighted the privacy awareness gap, where users are not fully aware of risks to their personal privacy resulting from their sharing information online (Keßler and McKenzie 2018, Coppens *et al.* 2014). In this paper, we consider the factors that contribute to privacy risks on GeoSNs and pay particular attention to users' awareness of their sharing behaviour when interacting on these networks.

Configuring (and updating) personal privacy settings on GeoSNs can be cumbersome, leading to possible divergence between users' sharing choices and previously specified sharing polices (Patil *et al.* 2014). Research is emerging that studies mechanisms for providing feedback to help raise user awareness of potential inconsistencies with default preferences (Tsai *et al.* 2009), which mainly relies on exposing how often they access privacy settings and encouraging them to revise their preferences actively. With the continuous accumulation of location tracks, constructing useable feedback becomes a challenge (Patil *et al.* 2014).

In this work we explore the problem of improving users' awareness of their location sharing behaviour and propose a model of privacy risks that is derived from studying the collective attitude of users towards sharing data on GeoSNs. The model can be used to detect vulnerable sharing scenarios and to inform the design of effective feedback notices in GeoSNs.

After analysing the types of information that are collected and mined on GeoSNs, we consider the following questions in the context of sharing location information.

(1) Does user awareness of the data they share and the possible processing or analysis that can be done over their data on GeoSNs affect their perception of personal

privacy? An experiment, involving 715 participants, is designed to gauge users' perception of risk when considering the data and its visibility to others in the context of different modes of users' awareness on GeoSNs.

(2) Can the users' perception of risk to personal privacy be modelled? The results from the experiment are used to devise a model of risk to personal privacy when sharing location information on GeoSNs. The model utilises users' willingness to share their data in different contexts to devise a measure of vulnerability in different sharing scenarios. A simple visualization of this vulnerability index is proposed to assist the user in understanding the level of risk to their personal privacy associated with their sharing behaviour.

The contribution of this work is twofold: first, we demonstrate that transparency of location data processing in GeoSNs can significantly impact users' perception of risk to their personal privacy; second, we propose a model of privacy risk on GeoSNs that is based on users' attitude to sharing their location. The paper starts in section 2 with a review of related works on user profiling from location tracks and users' privacy perception on online social networks and its implication. The dimensions of the location disclosure problem are examined in section 3. In section 4, the design of the experiment is presented and justified. Results that demonstrate the impact of the different contributing factors to privacy perception are analysed in section 5. In section 6, a model of risk to location privacy is proposed using the results obtained and a discussion of its utility within a privacy-aware GeoSN is presented, followed by conclusions in section 7.

2. Related work

An overview of research into user profiling using data collected from GeoSNs is given to highlight the range of information that can be extracted from this data. Studies on understanding users' perception of privacy on GeoSNs are reviewed, followed by an overview of current research efforts on designing privacy-aware systems.

2.1. *User Profiling on GeoSNs*

Understanding users from their location data collected on GeoSNs is an active area of research. Several studies considered the accurate identification of users' location from their GPS trails (Pontes *et al.* 2012, Bellatti *et al.* 2017). Using the user's profile of visited places and socio-historical ties, accurate prediction of future check-in information (Gao *et al.* 2012) and identification of user's home location (Gu *et al.* 2016) were demonstrated. Other works investigated the potential inference of social relationships between users of GeoSNs. For instance, users' co-occurrence in place, as extracted from geo-tagged Flickr photos, was sufficient for deducing, with high probability, the nature of their social ties and friendship links (Sadilek *et al.* 2012).

Recently, several research works examined the problem of extracting spatiotemporal movement and activity patterns of users on GeoSNs, for the purpose of understanding users and places. Mobility patterns on Foursquare were studied to identify popular places and to detect transition patterns between place categories (Noulas *et al.* 2011), while the distance between consecutive check-ins of users was used to compute their returning probability to venues (Preotiuc-Pietro and Cohn 2013). Kurashima et al. in (2013) demonstrated how geo-tagged content on Flickr can be used to understand

landmarks, topics of interest and active geographic regions of importance to the user and hence can recommend suitable travel routes.

With regards to understanding users, sensitive personal information can be revealed by tracking the user check-in information, including, gender, educational background, age and sexual orientation (Rossi and Musolesi 2014, Zhong *et al.* 2015). Liu et al. in (2018) summarise different modes of attack that can be used by adversaries on a mobile application to reveal the user's identity and to determine their position and time information, including machine learning methods (Murakami and Watanabe 2016) and collusion of malicious users. In this work, an adversary is any entity (person or organisation) that illegitimately (without the user's awareness or permission) seeks to collect user's data, whether for a useful purpose, e.g. making recommendations, or otherwise, e.g. stealing the user's identity. These can include the service provider, third parties or the user's friends.

2.2. *Privacy Perceptions on GeoSNs: The privacy Awareness Gap*

Early studies on Location-Based Services (LBS) showed that users were generally anxious about their privacy (Fisher *et al.* 2012) and will seek to manage it by deleting social connections, comments or by removing applications (Boyles and Smith 2012, Alrayes and Abdelmoty 2014). A study of Facebook users found that the amount of publicly displayed data decreases with time as users restrict the visibility of their profiles (Stutzman *et al.* 2013). Users tended to be more conservative with their sharing behaviour, selecting the most effective obfuscation methods, when they became aware of their location history

Although concerns about location privacy are evident, users will still share location information, driven in many cases by small rewards and incentives; a phenomenon known as the Privacy Paradox. An explanation of the inconsistency of privacy attitudes and privacy behaviour is an active area of research that requires in-depth study as noted in (Kokolakis 2017).

Methods to address the apparent gap between users' privacy awareness and the extent to which they share data are being proposed that try to assist users by learning their attitudes towards privacy. This can be achieved by directly asking users (Watson *et al.* 2015) or automatically by learning from the users' interaction behaviour and settings (Bilogrevic *et al.* 2016). Here we propose to model users' perceived privacy risks when disclosing location on GeoSNs and use this model within feedback tools to improve users' awareness.

2.3. *Privacy models and frameworks*

Privacy models provide principles and guidelines to be considered when designing a privacy-aware system. They present insights into how to design a system that serves users' awareness of potential privacy implications based on their interaction with it, and suggest means of effective privacy management by users. These models have common aspects, but can vary based on the application and privacy domains. A pioneering privacy model was introduced by Bellotti and Sellen in (1993), who proposed that the drivers of the design of a system should be the provision of feedback for and control of several aspects of information, including, information collection, processing, accessibility, and purpose of use. Their framework also identifies design criteria to help in evaluating design solutions. Adams and Sasse (1999) suggested three main

factors which help to define the boundaries under which a privacy breach can take place, namely, data sensitivity, information receiver, and information purpose of use. The importance of keeping the user informed of their information, disclosing actions through usable feedback, and providing the means to control privacy settings were advocated in several works (Langheinrich 2001, Friedman *et al.* 2005).

Shokri et al. in (2010) were the first to publish a unified framework of location privacy. Their review describes various location privacy preservation mechanisms (LPPMs) and compares metrics for measuring location privacy. The review was extended in (Liu *et al.* 2018) with a study of possible attack categories and location privacy metrics.

The above works examined how different aspects of location information can lead to potential privacy threats and some reviewed the efficacy of protection mechanisms to protect users against those threats. In this paper, we also consider the factors that contribute to privacy risks in GeoSNs but pay particular attention to users' awareness when interacting and sharing their information on these networks.

3. Dimensions of the Location Disclosure Problem

In this section we consider the factors that contribute to users' perception of risk to their personal privacy while interacting on GeoSNs. As the user's location footprints are accumulated over time, they become a rich source of information on the characteristics of the user as well as the places he visits. Whether this data is visible to others, and whether the user is aware of the extent of the information he is sharing, are factors that can influence his perception of personal privacy.

3.1. *The Data Dimensions*

The data dimensions comprise a group of three different dimensions that represent the different attributes of the data collected on GeoSNs, as follows.

(1) The spatial dimension (which places is the person visiting?)
(2) The socio-semantic dimension (what is the person doing in these places and with whom is he or she interacting?)
(3) The temporal dimension (when are these activities taking place?)

Separating the dimensions help to distinguish between physical location and behaviour, which may not always be linked in a directly observable way (for example, a user may be in a coffee shop, but working remotely rather than socialising). The type of data and the amount of data collected determine the kind of information that can be inferred and stored in the user profile. Hence, it is useful to study how the individual's perception of risk to personal privacy differ along these three dimensions.

(1) **The Spatial Dimension**
Presence of the user in a place is plotted on the spatial dimension. A track of user mobility in space is collected as a sequence of time-stamped geographic coordinates which can be reverse geo-coded to automatically detect the user's presence in specific places. In addition, the user may also indicate his presence in the place (e.g. by explicitly checking in). The latter case allows users to describe places of interest that are not digitised or identified on a general map.
(2) **The Socio-semantic Dimension**

This is a compound dimension and comprises two distinct aspects: a) explicit social links to other users, and b) shared content. Explicit links to other users, for example as friends or followers, is an orthogonal dimension to both the spatial and temporal dimensions. Here social ties are formed and maintained between users independently of their presence in geographic locations.

Shared content on GeoSNs refers to the different types of data provided by the users. This could include text (in a variety of forms such as tags, tips, reviews and tweets), images or videos. As is the case for social ties, content may be explicitly attached to a place visited, e.g. writing a tip when visiting a restaurant or tweeting about a music festival whilst attending it. Alternatively, the location may be independent of the shared content, e.g. tweeting about the release of a new album of a favourite artist whilst at home. Different semantic information concerning the user and their association with places can be extracted from the shared content. This could include the user's interests, activities and sentiments (Mohamed and Abdelmoty 2017).

(3) **The Temporal Dimension**

The temporal dimension gives a timeline of the user's visits to different places. The frequency of visits to geographic locations can be used as an indicator of the degree of association with the locations and with the related activities and concepts derived from the socio-semantic dimension. Clustering specific temporal intervals on the timeline can be made to study emerging patterns of user activity.

Regarding the spatial dimension, the sensitivity of a place is an important factor when considering privacy perception. Sensitivity of a place is an attribute that is linked to the type of information it reveals about the individual. For example, a hospital or a fertility clinic may be considered to be sensitive places since they are linked to a person's physical health. Other sensitive information that may be revealed from the place types include hobbies, religion and beliefs, political views, sexual orientation, physical or mental health, ethnic origin and commission of offence (as defined by the California Location Privacy Act of 2012 [1] and the Data Protection Act in the UK [2]). With regards to the social dimension, sharing information about being co-located with a particular person or a group of people may also be considered as sensitive information, since it may reveal the nature of the relationship between individuals. Also, sharing this information assumes an implied consent from the other people involved, which if disputed can amount to a potential privacy breach.

3.2. *The Visibility Dimension*

Visibility and/or accessibility of users' data will ultimately determine the level of threat to personal privacy, since if the data are not exposed or can't be accessed then there is no question of risk to personal privacy. Smith et al. in (2011) distinguished between two types of privacy: social privacy; which refers to an individuals' management of self-disclosure, accessibility to their information and availability to other people, and information privacy; which concerns controlling the accessibility to personal information, collection and exploitation by organisations and institutes. On GeoSNs, individuals are normally able to control social privacy by setting the visibility of their profile to either "Friends" (or "Followers") or "Public". In the former case, access to the individual's data is limited to a defined group of people, presumably known by this

[1] `https://www.eff.org/cases/california-location-privacy-act-2012` [Accessed: 14-May-2019]

[2] `http://www.legislation.gov.uk/ukpga/1998/29/section/2` [Accessed: 14-May-2019]

individual, while in the latter case, any user of the network can access the data. It is not normally possible to restrict the visibility of one's data for specific users, and thus all users within the "Friends" group have equal rights to access the individual's data irrespective of their degree of association to this individual. Information privacy, on the other hand, is usually determined by the terms, conditions and policies of handling data in the application. To conform with GDPR, GeoSNs need to ensure that their users are aware of what data are being collected and shared with third parties and how the information collected will be used. However, updating a privacy policy to include this information provides, in reality, a protection for the GeoSN against legal liability and is not itself sufficient to ensure a user is fully informed of how their information is processed whilst using the application.

For example, Facebook's recently updated data policy[3] indicate that they "collect information from and about the computers, phones, connected TVs and other web-connected devices you use that integrate with our Products, and we combine this information across different devices that you use .. information that we obtain from these devices includes: ..access to your GPS location, camera or photos". However, whilst using their Products, a user will not be aware of the amount of location tracks, places, events, activities, and other attributes that Facebook have collected (or derived) over time. Whilst users are able to restrict access to their precise device location, Facebook will collect the user location through their "IP addresses and information from your and others' use of Facebook Products (such as check-ins or events you attend)". Thus, in reality users' geo-profiling is done by default, though users living in the EU have the opportunity to object to the processing of their data, and if their objection is successful, have a right to request the erasure of their data under article 17(1)(c) of the GDPR.

Two levels of visibility are considered in this work: a) "Friends"; where the visibility of the user's profile is assumed to be restricted to a selected set of individuals or groups of individuals who are known to the user, and b) "Public"; where the user's profile is open and can be accessed by any other user of the GeoSN. The latter is a special case of information privacy where the data controllers of the GeoSN are considered as potential adversaries. Note that on *Facebook*, a third category of visibility is offered; namely "Friends of Friends". We estimated that creating a distinct category for this group would not be useful for our study and may confuse users, as they may not be able to distinguish the difference between the three groups when answering the questionnaire. For the purpose of generality and clarity, we have therefore considered this as a special case of the Friends category.

3.3. *The Awareness Dimension*

When interacting with any software system, user's attention is task oriented- they are aware of the task they are doing instantaneously. Their awareness of the data they are disclosing is bounded by the information needed for the task at hand. Privacy threats become apparent with accumulated information that can be mined from implicit relationships between data items over time. For example, when a user checks into a place at night, she may not be aware that this event can be used to deduce that this place is probably her home. This information can be derived by clustering multiple visits in the time dimension and analysing the frequency of visits. Awareness of the personal data shared and stored by the application is a critical factor to the user's perception

[3]https://en-gb.facebook.com/policy.php[Accessed: 14-May-2019]

Table 1.: Independent variables considered in the experiment and the corresponding study groups.

	Realistic		Attacker	
	Friends	Public	Friends	Public
Spatial	$S1_{R-F}$	$S2_{R-P}$	$S3_{A-F}$	$S4_{A-P}$
Spatial-Social	$SS1_{R-F}$	$SS2_{R-P}$	$SS3_{A-F}$	$SS4_{A-P}$
Spatial-Social-Temporal	$SST1_{R-F}$	$SST2_{R-P}$	$SST3_{A-F}$	$SST4_{A-P}$

of privacy. If the user is oblivious to the data stored in her profile, she will not be able to accurately perceive the potential risk to her privacy.

We attempt to study two modes of awareness: a) "Realistic" mode; this is the common mode of use of a GeoSN, where people are aware only of the data they are currently sharing and may also recall the visibility settings on their profile (i.e. whether their interaction is shared with a specific group of people), and b) "Attacker's" mode; that is where the GeoSN deliberately makes the user aware of not only the data itself, but also possible inferences that can possibly be derived by others who may have access to the data from their current interaction. Thus the difference between the two modes is the fact that one draws the user's attention to the implicit conclusions that can be drawn from their data, rather than simply the raw data itself. This increases the user's awareness of the privacy risks posed by a possible adversary. The latter case is hypothetical and is not supported by any major social network platforms currently on the market. It is envisioned as a possibility for a privacy-aware GeoSN that puts into practice the GDPR requirements of transparency and informed consent.

4. Experiment Design

The dimensions of the location disclosure problem above are used here to guide the design of a set of scenarios of use of a GeoSN. An experiment was carried out where participants were asked to consider the scenarios individually before deciding on their willingness to share their location. The participants' sharing decisions were then used to indicate how concerned they were about privacy when interacting on GeoSNs.

To understand the specific influence of the different aspects of the location-sharing problem, a between-subjects design[4] was adopted to examine the different study conditions, namely,

(1) *Data scope* (Spatial (S) vs Spatial-Social (SS) vs Spatial-Social-Temporal (SST))
(2) *Visibility scope* (Friends vs Public)
(3) *Awareness scope* (Realistic vs Attacker's).

Hence, to account for all combinations of the above, twelve treatment groups were needed in the experiment; four groups for each of the three data scopes, as shown in Table 1.

In each of the 12 groups, participants were asked to consider 10 scenarios of use of a GeoSN. The scopes of visibility and awareness were then used to frame a question which gauged their attitude to sharing location information. To understand the effect of place sensitivity on privacy concerns, different types of place were employed in

[4]Between-subjects design is an experiment where two or more groups of subjects are tested each by a different testing factor simultaneously (Wikipedia 2019).

Table 2.: Example Scenarios across the data dimensions with different types of places.

Scenarios	Spatial	Spatial-Social	Spatial-Social-Temporal
Insensitive place types	You are in a Mexican restaurant in town. You are having dinner with a friend. This is your first time in this place.	You are now in a Mexican restaurant in town. You have been here frequently in the past.	It is now Friday night and you are in the Village hall in your neighbourhood. You attend a drama group every Friday night in Spring.
Sensitive place types	You are in the Main Hospital in Town. You are there for your routine check-up. You visited this hospital only once in last year.	You are now in the Main Hospital in Town with Alex. You are both visiting a friend. You have visited this hospital only once last year.	You are now participating in a charity event in a religious centre (such as Church, Temple, Mosque, ...) that you belong to. You have regularly visited this place on Saturday afternoons in the last three months.
Personal places	You are now visiting your friend at 16 Park Place (an apartment building in town). You have not been here previously.	You are with Alex at your home at 16 Park Place (an apartment building). You have been here frequently with Alex in the past.	You are with Alex at your home at 16 Park Place (an apartment building). He normally visits on Sunday Evenings.

each of the 10 scenarios: 4 scenarios considered visiting public-insensitive place types (e.g. shopping mall, cinema, fitness centre, restaurant); 5 scenarios considered visiting public-sensitive place types (e.g. hospital, church, political party office, community centre for a particular ethnic group); and one scenario considered visiting a personal place (home). Table 2 shows some examples of the scenarios used in the different study groups.

Co-location with a friend was used across scenarios to represent variation on the social dimension. In particular, one of each scenario (sensitive and insensitive) was set as a visit with a close friend in the SS and SST groups. In the Spatial scenarios, no pattern of presence in a place is defined; visits are characterised as "unusual" or "occasional". In the Spatial-Social scenarios, a frequent pattern of presence was used to indicate a favourite activity or a close association to a friend. In the Spatial-Social-Temporal scenarios, regular presence, unusual visits or absence from a place were used to infer implicit temporal association with place as well as activity and social connection. Note that it would not have been realistic to isolate interactions on the spatial and temporal dimensions without considering the implication on the socio-semantic dimension. As a case in point consider the intrinsic link between the social habits of an individual and regular Tuesday visits to a Tennis club. It would not be possible to remove the semantic dimension from such a scenario since regularity often implies meaning of some kind - a regular personal or social activity. To control bias, the 10 scenarios were randomly presented to participants in every group. They were not linked to any particular real-world application. In the case of "Attacker's" scenarios, no reference to the purpose of using the data by the adversary was given. Participants were left to assume how their data might be used by others. t was clear from a post-study questionnaire, presented in Table A1 in the Appendix, that participants recognise the presence of adversaries online and that they value their online privacy.

In the *Friends* scenarios, participants were asked to answer the question: "Would you share your location now with your friends?". Similarly, in the *Public* scenarios, the question was: "Would you share your location now with other users?". In the *Attacker's* scenarios, the question was preceded by a statement to alert the user of the nature of information they are potentially revealing, as can be seen in the following examples (where 'Alex' was introduced as a close friend of the user on the social

network application).

- ($SS_{1_{R-F}}$) **:** You are now in the *Main Hospital* in Town *with Alex*. You are both visiting a friend. You have visited this hospital only once last year. Would you share your location now with friends?
- ($SST_{3_{A-F}}$)**:** It is now Saturday evening and you are in the Good Life Pub in town. You regularly go there on Weekends. If you share your location track, your *friend connections* will be able to see that you regularly go to this pub on Weekends. Would you share your location track now with friends?
- ($SST_{4_{A-F}}$) **:** It is now Saturday evening and you are in the Good Life Pub in town. You regularly go there on Weekends. If you share your location track, *other users of the application* will be able to see that you regularly go to this pub on Weekends. Would you share your location track now with other users?

Answers to the question are mapped to perception of risk to privacy; 'yes' corresponds to 'unconcerned', 'maybe' corresponds to 'concerned', and 'no' corresponds to 'very concerned'. To avoid bias, no direct mention of privacy is made in the wording of the questions, but participants were also able to justify their decision in an open-ended question after completing the scenarios. The majority of responses to this question were justifications directly related to privacy concerns (privacy, safety, sensitivity, protection, etc.) (86%), while the remaining 14% also mentioned other reasons such as social capital (what others think, interesting, useful, etc.) [5]. It is noted that the social capital concerns featured mainly in the spatial scenarios and concerns became more privacy-oriented as more information was revealed in the scenarios.

4.1. *Participants and Procedure*

Participants were recruited using Amazon Mechanical Turk (MTurk)[6]. To ensure that participants were able to relate to the presented scenarios, a qualification test for the study was used to choose experienced MTurk workers with good reputation (have $\geq$ 95% approval rate for at least 500 tasks on MTurk) and who share their location from their social network accounts. Constraints were also enforced to limit participation in the study to only one time. The dissemination of the different versions of the study was carried out at different times throughout the day to enable participation from eligible workers from any country. The order of the scenarios was randomly presented to every participant in all treatment groups.

747 participants entered the study, 32 of whom did not meet the criteria of sharing location information on GeoSNs. The remaining 715 participants were able to complete the survey in an average of 6.14 minutes. The number of participants in each treatment group is shown in Table 3. The sample was young (Mean= 33.35 years old, SD= 9.88) with an equal distribution of males and females. Most of the participants were from North America (72.31%), with significant representation in Asia (18.04%) and Europe (6.15%). The majority of participants use social network applications frequently (several times a day) (69.79%). Facebook was the most used platform to share location information, followed by Twitter, Instagram and Google+. Users of these applications represented 95%, 55%, 53% and 41% of participants respectively. Participants also tag

[5]The survey and all the responses to this question can be accessed at: `http://doi.org/10.17035/d.2019.0075767525`

[6]MTurk is a widely used online crowdsourcing platform for virtually leveraging a distributed workforce for tasks requiring human input, such as survey participation, and is used in similar studies (Rader 2014).

Table 3.: Number of participants in each of the study groups.

Spatial		**Spatial-Social**		**Spatial-Social-Temporal**	
$S1_{R-F}$	59	$SS1_{R-F}$	60	$SST1_{R-F}$	58
$S2_{R-P}$	59	$SS2_{R-P}$	60	$SST2_{R-P}$	59
$S3_{A-F}$	58	$SS3_{A-F}$	59	$SST3_{A-F}$	62
$S4_{A-P}$	59	$SS4_{A-P}$	61	$SST4_{A-P}$	61

Table 4.: Average sharing decisions for participants in all study groups.

Spatial				**Spatial-Social**				**Spatial-Social-Temporal**			
	Yes	**Maybe**	**No**		**Yes**	**Maybe**	**No**		**Yes**	**Maybe**	**No**
$S1_{R-F}$	*56.10%*	19.49%	24.41%	$SS1_{R-F}$	47.65%	26.74%	25.61%	$SST1_{R-F}$	50.09%	34.61%	28.23%
$S2_{R-P}$	45.25%	27.29%	27.46%	$SS2_{R-P}$	40.56%	26.11%	33.33%	$SST2_{R-P}$	41.17%	29.87%	36.43%
$S3_{A-F}$	42.59%	24.83%	32.59%	$SS3_{A-F}$	46.67%	26.85%	26.48%	$SST3_{A-F}$	44.81%	30.60%	30.24%
$S4_{A-P}$	40.34%	28.98%	30.68%	$SS4_{A-P}$	36.07%	30.97%	32.97%	$SST4_{A-P}$	37.70%	23.13%	*50.27%*

their friends when sharing location information (always: 15.66%, sometimes: 78.74%). 22.66% of participants enable *location services* or other similar location features on mobile applications frequently (always on) and 70.35% enable them moderately (when required by an application), while only 4.48% disable such features.

A pilot study was carried out with five postgraduate students, who were tasked with completing different versions of the study. The study and the scenarios were perceived as easy to understand and follow. Feedback given in the post-study interview was mainly related to improving the wording of some scenarios.

5. Results

The study involved three independent/predictor variables representing the study conditions (data dimensions (Spatial, Spatial-Social or Spatial-Social-Temporal), awareness (Realistic or Attacker's), and visibility (Friends or Public)) and one dependent/outcome variable representing the participants' location-sharing decision (yes, maybe or no). Table 4 summarises the sharing decisions among participants in each of the study groups.

A Chi-square test of independence was used to examine the impact of the study conditions on the participants' attitude to privacy. Spearman's Rank-Order Correlation was also used to examine the strength and direction of the correlation (if any) between the study conditions and participant's perceptions. An ordinal logistic regression model was adopted where the levels of the outcome variables were coded as follows: Yes=1, Maybe=2, and No=3. To interpret the regression results, positive coefficients (>0) were noted to indicate a greater likelihood of willingness to share location (i.e. not being concerned); coefficients equal to 0 were used to indicate no additional likelihood on top of the baseline, and negative coefficients (<0) were used to indicate a lower likelihood of willingness to share (higher likelihood of being concerned). The results of the model are shown in Table 5, where it can be seen that compared to the Spatial-Social-Temporal scenarios, participants were more willing to share their location in the Spatial-Social scenarios and even more so in the Spatial scenarios. On the other hand, participants were less likely to share their location in the Attacker's scenarios compared to the Realistic scenarios and in Public scenarios compared to

Table 5.: Results of ordinal logistic regression model examining the impact of the study conditions on the participant's attitude to privacy.

Condition	**Estimate**	**Odds ratio**	**Std. Error**	**P(Sig.)**	**95% Confidence Interval**	
					Lower Bound	**Upper Bound**
Data Dimensions (baseline= Spatial-Social-Temporal)						
Dimension=Spatial	.252	1.287	.054	<.0001	.147	.358
Dimension=Spatial-Social	.151	1.163	.055	.006	.043	.259
Visibility (baseline=Public)						
Visibility=Friends	.320	1.378	.045	<.0001	.233	.408
Awareness (baseline=Realistic)						
Awareness=Attackers' View	-.217	.805	.045	<.0001	-.305	-.129

Friends.

All the study conditions; data dimensions, visibility and awareness of location-sharing activities were shown to significantly impact the participants' privacy perceptions (p<.0001). In particular, visibility was the factor with the strongest impact on privacy perception followed the awareness and data scopes. It is interesting to note that place sensitivity and co-location with a friend have also been shown to significantly influence the participants' privacy attitude. The sensitivity of place reduced the participants' willingness to share by 31% while co-location with a friend reduced it by 8%. A more detailed analysis of the results is given below.

5.1. *Impact of the Data Dimensions*

The data dimensions were shown to have a statistically significant impact on users' willingness to share their location (Pearson Chi-Square= 22.72, p<0.0001). A moderately positive correlation between the data dimensions and the participants' attitude to sharing their information was noted (Spearman's rho=0.53, p<0.0001). Hence, users tend to become more concerned about their location privacy as the information shared becomes more complex along the different data dimensions. Participants in the *Spatial* study groups were the least concerned about their privacy (maybe 25.1%:, no: 28.8%), compared to the *Spatial-Social* groups (maybe 27.7%:, no: 29.6%) and the *Spatial-Social-Temporal* groups (maybe 27%:, no: 33.2%).

Responses to the open-ended question revealed increasingly more privacy as well as safety concerns as more data dimensions are revealed. Example responses from the *Spatial-Social-Temporal* groups refer to fears of tracking by others: the fear that "*someone... could track you and get to you if they wanted to*" and being "*concerned about... safety. Some people could see the pattern of my whereabouts, and use that information to stalk me or my friend*"."*Because there are some places you just do not need to let others know [the location of]... These days people could try to come to your home and rape you, murder you, or even kidnap you*". Some responses showed awareness of absence inference from location tracks, e.g. "*Sharing my location... on a regular basis advertises that I am not at home on those days and times*" and "*I would not let anyone know where I go on a regular basis. This is a good way to have your home robbed*".

5.1.1. *Impact of Place Sensitivity*

Observing the sensitivity of place across all the data dimensions has shown that it has a statistically significant effect on users' willingness to share their information (Pearson Chi-Square= 729.903, p<0.0001). As can be expected, participants were most likely to share their location in public-insensitive place types (Yes: 58.4%). Their willingness to share decreases significantly, by 31%, in personal places (Yes: 27.4%) and to a large degree (by 25%) in other sensitive place types (Yes: 33.2%).

Reluctance to share personal location in sensitive places was explained in responses to open-ended questions, for example, it was said that "*Sharing location information for public places is mostly OK... but sharing personal location information related to religion, political affiliation or a friend's house via location info is something I try not to do*" and that one "*wouldn't want to share medical location places or anything having to do with my culture, faith or home. Those are private issues*". Figure 1 shows the sharing decisions across the sensitive and personal place types.

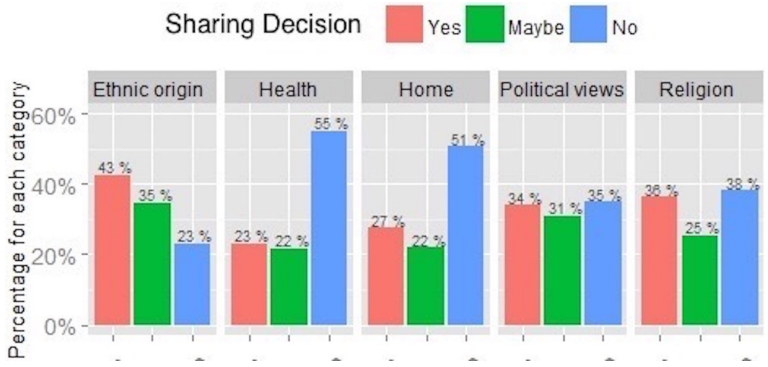


Figure 1.: Sharing decisions in sensitive and personal places.

5.1.2. *Impact of co-location with Friends*

Co-location with a friend was also shown to have a statistically significant effect on the participants' willingness to share their location (Pearson Chi-Square= 46.363, p<0.0001). Participants were less likely to share their location if they are with a friend than when being at a place by themselves (Yes = 46% when alone compared to 38% when with a friend).

Some reasons for the sharing attitude are explained by participants to be mainly due to considering the information as sensitive or that it involves someone else whose privacy should be considered, as shown in the following comments: "*To protect the privacy of other people I was with or visiting*" and "*if I do tag friends, I like to ask permission from them first*".

5.2. *Impact of the Visibility Scope*

The visibility scope of the user profile in location-sharing scenarios has a statistically significant impact on the users' likelihood to disclose their location (Pearson Chi-Square= 50.204, p<0.0001). A strong positive correlation is noted between the participants' privacy attitude and the visibility of their information (Spearman's rho=0.85, p<0.0001). This suggests that participants are less likely to share their location if their profile was Public than if it was set to be visible by Friends only, as indeed confirmed

in the results (34% said No to sharing with Public compared to 27% with Friends). Participants who were reluctant to disclose their location with other users justified their attitude by referring to their desire to protect their privacy, e.g. "*I didn't want to disclose my location for strangers to know*", "*I don't want people to know where I am or potentially stalk me" and "Some occasions seem too personal to share with the public*".

Figure 2 shows the combined results for the data dimensions and the visibility. As shown in the figure, participants are most willing to share their locations in the S_F scenarios (Yes: 49.34%), while they are mostly unwilling to share their location in the SST_P scenarios (No: 40%). The impact of the visibility scope is evident in the figure, where the difference in the sharing decisions is more pronounced in the Public scenarios, as was indicated in the participants' responses: "*I am not comfortable with strangers having access to my address and access to my routines*".

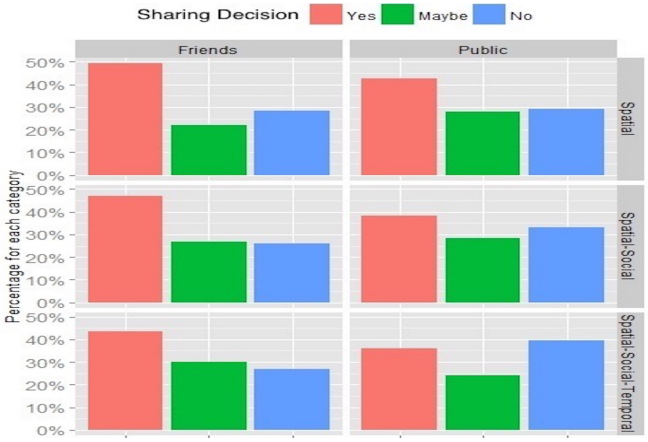


Figure 2.: Sharing decisions categorised by visibility scope and data dimensions.

5.3. *Impact of Awareness*

Users' awareness of implications of their sharing decisions has a statistically significant impact on the likelihood of them disclosing their location (Pearson Chi-Square= 23.340, $p<0.0001$). A fairly strong positive correlation is noted between the participants' privacy attitude and their awareness (Spearman's rho=0.58, $p<0.0001$). This suggests that participants are less likely to share their location if they were made aware of the nature of their disclosed information and its possible implications.

Justification for the sharing decision highlighted users' awareness of the privacy implications and their need to control their privacy, e.g. "*I wouldn't want to broadcast my history of the place*", "*I didn't want to be tracked in sensitive areas*" and "*I don't want to be tracked and I don't want someone to notice patterns [in] where I go*". Being mindful of the sensitivity of the places visited triggered a reaction to question their sharing behaviour (often due to users' personal affiliations with particular types of place); "*I prefer to keep certain things private - politics, health info, and any other information that could be used to deduce other things I prefer to keep private*", "*No one needs to know I go to church on Sundays for example*", and "*Sharing some locations would allow other users (who I may not want to share that data with) to interpret or assume things about me that I would not necessarily want to be public knowledge*".

Figure 3(a) shows the effect of the awareness factor grouped by the data dimension,

and in (b) the effect is grouped by the visibility scope. Participants are most willing to share their locations in the S_R groups (Yes:50.68%) and are least willing to share their data in the SST_A groups (No: 37%). The impact of the visibility scope is evident in Figure 3, where the differences in the sharing decisions are more pronounced in the Public scenarios, as was indicated in the participants' responses: "*I am not comfortable with strangers having access to my address and access to my routines*".

(a) (b)

Figure 3.: Sharing decisions grouped by a) data dimensions and awareness conditions and b) visibility scope and awareness conditions.

Exploration of the relationships between all study conditions in terms of their mutual impact on users' privacy perceptions reveals that the participants most likely to share their location are those presented with S_{R-F} scenarios (Yes: 56.10%), followed by those in SST_{R-F} and SS_{R-F} (Yes:50% and 48% respectively). This observation suggests that participants were least concerned about their location privacy when sharing with friends in realistic experience and the extracted information is at minimum (data dimension=Spatial), and their concern increases when more personal information are revealed. On the contrary, participants were least willing to disclose their location in SST_{A-P} scenarios (No: 50.27%), followed by SS_{A-P} and S_{A-P} scenarios (No: 33% and 31% respectively).

5.4. *Discussion*

The validity of this study was carefully considered. The use of hypothetical location-sharing scenarios has been shown to be an effective approach for yielding generalisable outcomes in a number of previous studies (e.g. (Patil *et al.* 2014, Tang *et al.* 2011)). Using this approach has the advantage of removing the association and dependence on a specific GeoSN and hence reducing the effects of particular interface and interaction modes offered within those applications, whilst also offering the opportunity of administration to large samples of users. Bias was limited by careful choice of participants, using a between-subjects design and the random order of presentation of the scenarios in all study groups.

The scenarios were developed in a consistent manner to cover all the variables of interest, whilst at the same time enabling the participants to be immersed to a large degree in the location-sharing experience. This was evident in the participants' responses to the open-ended questions. For example, they would refer to themselves: "*When I was in a political meeting, I would not share my location without the permission of others or the party*" and "*I answered maybe or no to places that I frequent because someone might see a pattern*". They also referred to personal experiences: "*I*

have a friend that checked in everywhere she went. Her profile was public so anyone could see it. Her house was burglarized twice in one month because everyone knew she wasn't home." and "*I started to think about how if it was a regular routine that I was somewhere on a specific night, it was going to give other people potentially too much information about my habits and whereabouts*".

We collected demographics data on age, gender and nationality only. An analysis of the effect of these attributes on the sharing decisions is given in the Appendix. The trends noted there may prove useful in the design of future studies. Several points regarding the impact of personal characteristics on sharing decisions are noted below, that would also be useful to consider in more detail in future work.

(1) It can be argued that people's attitude to location privacy is unique and depends on personal characteristics, such as religious belief or political affiliation. For example, a religious person may not wish to expose her presence in a place of worship. A person living in a city may have a stronger sense of location privacy compared to someone living in a small village in the countryside. This argument was supported to some degree in the responses to the open-ended question reported above. Ideally, we could have tailored the scenarios related to some of the sensitive place types to the particular characteristics of the participants. Such personalised tailoring of the questions was difficult to perform in a widely disseminated online study such as this. A more in-depth study is needed in the future that can control and measure the effect of such personal attributes on the perception of risk to location privacy.
(2) Similarly, culture may play a role in people's attitude to online interaction. This study didn't probe or limit the scope of the origin of participants beyond recording their nationality, although several participants made reference to culturally specific privacy concerns such as religion in the open-ended responses.
(3) The study didn't control for the factor of prior adverse experience with privacy (Trepte *et al.* 2014). The importance of this factor is becoming evident as more people become involved or become aware of negative privacy experiences. A more in-depth study of the effects of this factor is needed in future work.

In what follows the results of the experiment are used to guide a model of perception of risk to personal privacy on GeoSNs.

6. A Model of Privacy Risks of Location Sharing on GeoSNs

As has been found in previous studies (Kumaraguru and Cranor 2005), a large proportion of people can be considered to be privacy pragmatists who are willing to share their data if they see tangible benefits for doing so, but are also very concerned about protecting themselves from the abuse and misuse of their personal information. In this section, we use the results of the analysis above to inform the derivation of a model of user's perception of risk when sharing their location on GeoSNs. To facilitate its comprehension, a traffic light metaphor is used here to represent the level of risk: Green = Unconcerned, Amber = Concerned and Red = Very concerned. We can map this classification to the Westin and Harris privacy scale (Kumaraguru and Cranor 2005): Green = Privacy unconcerned, Amber = Privacy Pragmatists and Red = Privacy Fundamentalist and observe how they correspond to one another. A colour-coded privacy notice design is also desirable due to its simplicity and universality. The colours will range from red (the information is revealing and may not be safe to disclose),

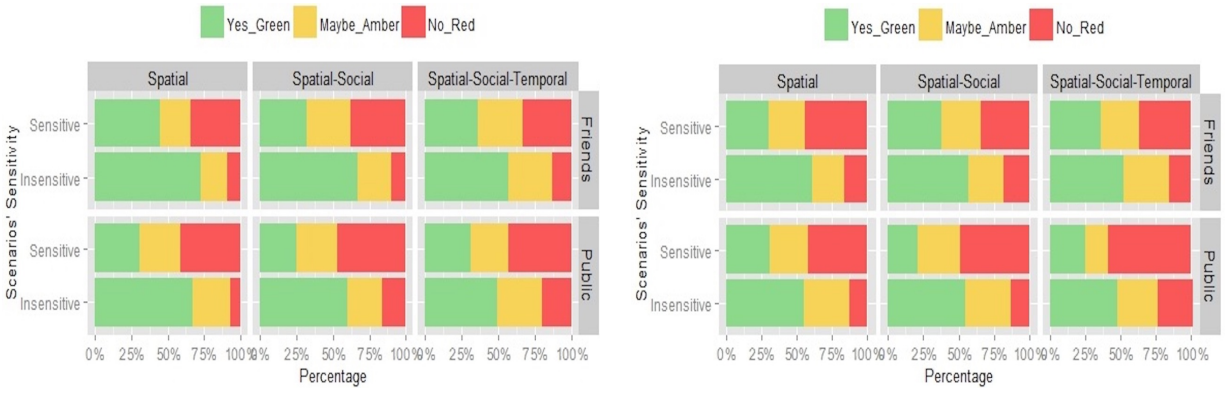
Table 6.: Sharing decisions classified by the sensitivity of place types in the Realistic awareness scenarios.

Visibility		Friends			Public		
Data Dimension	**Sensitivity**	*Yes*	*Maybe*	*No*	*Yes*	*Maybe*	*No*
Spatial	Insensitive	73%	18%	9%	67%	26%	7%
	Sensitive	45%	21%	34%	31%	28%	41%
Spatial-Social	Insensitive	67%	23%	10%	60%	24%	16%
	Sensitive	32%	30%	38%	25%	28%	47%
Spatial-Social-Temporal	Insensitive	57%	30%	13%	49%	31%	20%
	Sensitive	36%	31%	33%	31%	26%	43%

Table 7.: Sharing decisions classified by the sensitivity of place types in the Attacker's awareness scenarios.

Visibility		Friends			Public		
Data Dimension	**Sensitivity**	*Yes*	*Maybe*	*No*	*Yes*	*Maybe*	*No*
Spatial	Insensitive	61%	23%	16%	55%	32%	13%
	Sensitive	30%	26%	44%	31%	27%	42%
Spatial-Social	Insensitive	57%	25%	18%	55%	32%	13%
	Sensitive	38%	28%	34%	21%	30%	49%
Spatial-Social-Temporal	Insensitive	52%	32%	16%	48%	28%	25%
	Sensitive	36%	27%	37%	25%	16%	59%

through to amber (one may need to exercise caution when disclosing this information) and ultimately green (information that may be safely disclosed). Participants' sharing decisions based on the variables used in the study are presented in Table 6 for the Realistic scenarios and in Table 7 for the Attacker's scenarios. Figure 4 shows the mapping of this data using the traffic light metaphor in both states of awareness.



(a) (b)

Figure 4.: A mapping of sharing decisions to three privacy risk levels in the case of (a)Realistic and (b) Attacker's mode of awareness.

Given the responses, we wish to produce a model which bases its decisions on the shareability of data on user feedback. One way of doing this is to produce a relative scale where the upper and lower limits are defined by the upper and lower limits of the data itself. In Table 8, the responses were reclassified to only two categories 'Yes' and 'No'. The 'Maybe' responses were evenly distributed between these two categories. The threshold values were computed by considering the most vulnerable and least vulnerable privacy groups according to the revised percentage of 'Yes' Responses.

The situation with the largest percentage of Yes responses is Spa-

Table 8.: Sharing decisions with only two groups of responses; 'Yes' and 'No'.

Awareness		*Realistic*				*Attacker's view*			
Visibility		*Friends*		*Public*		*Friends*		*Public*	
Data Dimensions	**Sensitivity**	Yes	No	Yes	No	Yes	No	Yes	No
Spatial	*Insensitive*	**82%**	18%	80%	20%	72.5%	27.5%	71%	29%
	Sensitive	55.5%	44.5%	45%	55%	43%	57%	44.5%	55.5%
Spatial-Social	*Insensitive*	78.5%	21.5%	72%	28%	69.5%	30.5%	71%	29%
	Sensitive	47%	53%	39%	61%	52%	48%,	36%	64%
Spatial-Social -Temporal	*Insensitive*	72%	28%	64.5%	35.5%	68%	26%	62%	39%
	Sensitive	51 %	49%	44%	56%	49.5%	50.5%	33%	**67%**

tial:Insensitive:Friends:Realistic as seen in Table 8, with Yes = 82%. On the other hand, the situation with the smallest percentage of Yes responses (equivalently, the largest percentage of No responses) was Spatial-Social-Temporal:Sensitive:Public:Attacker's with Yes = 33% (both highlighted in Table 8). Semantically, we can take these groups to represent the "safest" and "least safe" situations as judged by the respondents.

We propose a so-called Vulnerability Index (VI) which will be defined as a relative percentage, using the above percentages as upper and lower bounds respectively. A Yes percentage at or above the upper threshold of Yes = 82% will be capped at a VI of 100%, and a percentage equal to or below the lower threshold of Yes = 33% will be capped at a VI of 0%. Mathematically, the function is as follows, where x is the percentage of Yes responses:

$$VI(x) = \begin{cases} 100, & 82 < x \leqslant 100 \\ F(x), & 33 \leqslant x \leqslant 82 \\ 0, & 0 \leqslant x < 33 \end{cases}.$$

We note that $F(x)$ might be any function and its shape will determine the nature of our privacy recommendations. For example, by biasing the function against lower values, we can be more conservative in our recommendations. As a starting point and a useful illustration, a simple, linear $F(x)$ might be as follows:

$$F(x) = \frac{x - 33}{82 - 33} \times 100 = \frac{x - 33}{49} \times 100 \tag{1}$$

The output of the above function $VI(x)$, with $F(x)$ as in Equation 1, can be interpreted as the degree to which a particular situation's responses tend towards one of the measured extremes.

To generate a traffic light colour, we used the RGB colour scheme, which composes a colour from three colour values: Red, Green and Blue. The formulae below are designed to give a linear variation in colour between entirely Red at $VI = 0 : RGB(255, 0, 0)$; and entirely Green at $VI = 100 : RGB(0, 255, 0)$. At $VI = 50$, we have equal measures of Red and Green giving yellow: $RGB(255, 255, 0)$.

$$R = \begin{cases} 255, & 0 < VI \leqslant 50 \\ (1 - \frac{VI - 50}{50}) \times 255, & 50 \leqslant VI \leqslant 100 \end{cases}.$$

$$G = \begin{cases} 255, & 50 < VI \leqslant 100 \\ (1 - \frac{50 - VI}{50}) \times 255, & 0 \leqslant VI \leqslant 50 \end{cases}.$$

Table 9 uses the traffic light model to present the data from Table 8 with the original Yes responses, while Table 10 presents the results using the Vulnerability Indexes. Table 11 presents a different view of the model that combines both the Realistic and Attacker's views and assigns definite risk levels to all groups according to the average of the Vulnerability Index along the Awareness dimension. The average Vulnerability

Table 9.: A proposed model of risk levels based on the proportion of sharing decisions. The percentage of Yes responses is given. The colour is generated from the Vulnerability Index, ranging from red to green.

Awareness		*Realistic*		*Attacker's view*	
Visibility		*Friends*	*Public*	*Friends*	*Public*
Data Dimension	**Sensitivity**				
Spatial	*Insensitive*	82	80	72.5	71
	Sensitive	55.5	45	43	44.5
Spatial-Social	*Insensitive*	78.5	72	69.5	71
	Sensitive	47	39	52	36
Spatial-Social -Temporal	*Insensitive*	72	64.5	68	62
	Sensitive	51	44	49.5	33

Table 10.: The Vulnerability Indexes for each scenario. The associated traffic light colours ranging from red to green are shown.

Awareness		*Realistic*		*Attacker's view*	
Visibility		*Friends*	*Public*	*Friends*	*Public*
Data Dimension	**Sensitivity**				
Spatial	*Insensitive*	100	95.92	80.61	77.55
	Sensitive	45.91	24.49	20.41	23.47
Spatial-Social	*Insensitive*	92.86	79.59	74.49	77.55
	Sensitive	28.57	12.24	38.78	6.12
Spatial-Social -Temporal	*Insensitive*	79.59	64.29	71.43	59.18
	Sensitive	36.73	22.45	33.67	0

Table 11.: A revised risk model with colours dictated by the Vulnerability Indexes in Table 9 averaged along the Awareness dimension.

	Friends		**Public**	
	Insensitive	***Sensitive***	***Insensitive***	***Sensitive***
Spatial	90.31	33.16	86.73	23.98
Spatial-Social	83.67	33.67	78.57	9.18
Spatial-Social-Temporal	75.51	35.20	61.73	11.22

Index is then shown shaded in the corresponding traffic light colour. The average Vulnerability Index (VI) over all dimensions (data, visibility and awareness) was 51.91. This is almost exactly halfway between the defined maximum (100) and the defined minimum (0). This is an indication that our linear $F(x)$ used for interpolation between the maximum and the minimum (shown in Equation 1) results in a Vulnerability Index with values which are evenly spread over the scale. A VI of 51.91 corresponds to an amber traffic light colour, which when mapped to the Westin and Harris scale indicates that users of GeoSNs are on average privacy pragmatists who are willing to share their location information, but need to be supported by the applications to ensure that their data are protected from misuse or abuse.

Table 11 presents a possible model of interpreting users' perception of privacy risk when disclosing location information. The model can be used by GeoSNs to improve users' awareness of their sharing behaviour. Based on the information collected in the user profiles, the application can predict a level of risk for future sharing decisions and alert the user as appropriate. The colour can be used to indicate to the user in a simple manner the extent to which the item they are sharing poses a privacy risk. The visual interpretation in the form of traffic light colours is just one such method. Another simple example would be to map the VI to a point on a scale from 1 to 10. Privacy perception is ultimately a personal variable, however, the above derived risk

models present a baseline; an indication of sharing behaviour of majority of users on the GeoSN, against which individual user's behaviour can be compared. Ultimately, the application would be able to learn from the individual's behaviour and adjust the model to suit.

From both tables, it can be seen that the sensitivity of place plays a major role in perception of risk on GeoSNs, where users are mostly willing to share their location when visiting insensitive places. A simple method of enhancing privacy in GeoSNs can therefore focus on sensitive places for users. These places can be identified by the user manually or can be automatically identified from their location tracks. The application can then alert the user mainly when sharing information whilst visiting those places.

7. Conclusions

This paper presented a data-oriented approach to understanding users' perception of threat to privacy on GeoSNs and proposes a model of privacy risks that is derived from studying collective users' attitude to sharing data on GeoSNs. Aspects of the problem have been identified, namely, data, visibility and awareness. Data disclosed by sharing location information vary within a space defined by the spatial, temporal and socio-semantic dimensions. In addition, the sensitivity of the places visited as well as co-location with others were identified as contributing factors to privacy risks. An experiment has been designed to assess the effect of these variables on the privacy perception of GeoSN users. Perception of risk was noted to increase as the information content in the data disclosed increased; whether with the data dimension, or with place sensitivity and co-location with other people. Visibility of the information was shown to have the most significant impact on privacy perception, where users were more comfortable sharing their information with 'Friends' (as defined by connections made on social web applications). Making the user aware of the nature of the information they are sharing was seen to have a significant impact on their sharing behaviour. This indicates that users' awareness is limited when interacting on GeoSNs and thus questions their presumed consent of use of the applications.

The study involved 715 participants split into 12 groups to study the different identified variables. Results from the study were used to define a simple model of privacy risk on GeoSNs. The model shows that in the majority of cases, users can be considered privacy pragmatists who are willing to share their data if they see tangible benefits for doing so, but are also very concerned about protecting themselves from the abuse and misuse of their personal information. GeoSNs can employ such a model to design privacy notification systems to alert the users to possible consequences of sharing information, thus allowing the user to take control of their sharing behaviour and ultimately increasing the trust in the application. These privacy notifications can be deduced from relative measures, such as the proposed Vulnerability Index, and visualised in ways which are easy to understand, for example a traffic light scale.

Future work will consider the utility of the proposed model in designing privacy-aware GeoSNs and the possible effects of different models of connection on online social networks, e.g. Friends of Friends. Usability of the designs will need to be considered along with a study of cost-benefit analysis of employing such models in GeoSNs.

References

Adams, A. and Sasse, M.A., 1999. Privacy issues in ubiquitous multimedia environments: Wake sleeping dogs, or let them lie. *In*: *Proceedings of INTERACT*. vol. 99, 214–221.

Alrayes, F. and Abdelmoty, A., 2014. Privacy concerns due to location sharing on geo-social networks. *International Journal On Advances in Security*, 7 (3 and 4), 62–75.

BBC, 2018. Facebook scandal 'hit 87 million users'. `https://www.bbc.co.uk/news/technology-43649018#`, June.

Bellatti, J., *et al.*, 2017. Driving habits data: Location privacy implications and solutions. *IEEE Security and Privacy*, 15 (1), 12–20.

Bellotti, V. and Sellen, A., 1993. Design for privacy in ubiquitous computing environments. *In*: *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93*. Springer, 77–92.

Bilogrevic, I., *et al.*, 2016. A machine-learning based approach to privacy-aware information-sharing in mobile social networks. *Pervasive Mobile Computing*, 25, 125–142.

Boyles, J. and Smith, A., 2012. Privacy and data management on mobile devices. *Pew Internet & American Life Project*.

Coppens, P., *et al.*, 2014. Privacy in location-based social networks: Researching the interrelatedness of scripts and usage. *In*: *Proceedings of the Symposium on Usable Privacy and Security*.

EUR-Lex, 2018. Regulation (eu) 2016/679 of the european parliament. `https://eur-lex.europa.eu/eli/reg/2016/679/oj`, June.

Fisher, D., Dorner, L., and Wagner, D., 2012. Location privacy: user behavior in the field. *In*: *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*. ACM, 51–56.

Friedman, B., Lin, P., and Miller, J., 2005. Informed consent by design. *Security and Usability*, 495–521.

Gao, H., Tang, J., and Liu, H., 2012. gSCorr: modeling geo-social correlations for new check-ins on location-based social networks. *In*: *Proceedings of the 21st ACM international Conference on Information and Knowledge Management, CIKM '12*. 1582–1586.

Gu, Y., *et al.*, 2016. We know where you are: Home location identification in location-based social networks. *In*: *Computer Communication and Networks (ICCCN), 25th International Conference on*. IEEE, 1–9.

Keßler, C. and McKenzie, G., 2018. A geoprivacy manifesto. *Transactions in GIS*, 22, 3–19.

Kokolakis, S., 2017. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122–134.

Kumaraguru, P. and Cranor, L., 2005. *Privacy indexes: a survey of westin studies*. Institue for Software Research, Carnegie Mellon University, 1-22.

Kurashima, T., *et al.*, 2013. Geo topic model: joint modeling of user's activity area and interests for location recommendation. *In*: *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 375–384.

Langheinrich, M., 2001. Privacy by design: Principles of privacy-aware ubiquitous systems. *In*: *Ubicomp 2001: Ubiquitous Computing*. Springer, 273–291.

Liu, B., *et al.*, 2018. Location Privacy and Its Applications: A Systematic Study. *IEEE Access*, 6, 17606 – 17624.

Mohamed, S. and Abdelmoty, A., 2017. Spatio-semantic user profiles in location-based social networks. *International Journal of Data Science and Analytics*, 4 (2), 127–142.

Murakami, T. and Watanabe, H., 2016. Localization attacks using matrix and tensor factorization. *IEEE Trans. Inf. Forensics Security*, 11 (8), 1647–1660.

Noulas, A., *et al.*, 2011. An empirical study of geographic user activity patterns in foursquare. *In*: *ICWSM*. 70–73.

Patil, S., *et al.*, 2014. Reflection or action? how feedback and control affect location sharing decisions. *In*: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 101–110.

Pontes, T., *et al.*, 2012. We know where you live?: privacy characterization of foursquare behavior. *In*: *UbiComp '12 Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 898–905.

Preotiuc-Pietro, D. and Cohn, T., 2013. Mining user behaviours: a study of check-in patterns in location based social networks. *In*: *Proceedings of the 5th Annual ACM Web Science Conference*. 306–315.

Rader, E., 2014. Awareness of behavioral tracking and information privacy concern in facebook and google. *In*: *Proc. of Symposium on Usable Privacy and Security (SOUPS), Menlo Park, CA, USA*. ACM, 51–68.

Rossi, L. and Musolesi, M., 2014. It's the way you check-in: identifying users in location-based social networks. *In*: *Proceedings of the second edition of the ACM conference on Online social networks*. ACM, 215–226.

Sadilek, A., Kautz, H., and Bigham, J., 2012. Finding your friends and following them to where you are. *In*: *Proceedings of the fifth ACM international conference on Web Search and Data Mining, WSDM '12*. 723–732.

Shokri, R., Freudiger, J., and Hubaux, J.P., 2010. *A unified framework for location privacy*. EPFL.

Smith, H., Dinev, T., and Xu, H., 2011. Theory and review information privacy research: An interdisciplinary review 1. *MISQ Quarterly*, 35 (4), 989–1015.

Stutzman, F., Gross, R., and Acquisti, A., 2013. Silent listeners: The evolution of privacy and disclosure on facebook. *Journal of privacy and confidentiality*, 4 (2), 7–41.

Tang, K., Hong, J., and Siewiorek, D., 2011. Understanding how visual representations of location feeds affect end-user privacy concerns. *In*: *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 207–216.

Trepte, S., Dienlin, T., and Reinecke, L., 2014. Risky behaviors: How online experiences influence privacy behaviors. *In*: B. Stark, O. Quiring and N. Jackob, eds. *From the gutenberg galaxy to the google galaxy*. UVK, 225–244.

Tsai, J., *et al.*, 2009. Who's viewed you? the impact of feedback in a mobile location-sharing application. *In*: *CHI '09, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2003–2012.

Watson, J., Lipford, H., and Besmer, A., 2015. Mapping user preference to privacy default settings. *ACM Trans. Comput.-Human Interact.*, 22 (6), 32:1–32:20.

Wikipedia, 2019. Between-group design. `https://en.wikipedia.org/wiki/Between-group_design`, Mayi.

Zhong, Y., *et al.*, 2015. You are where you go: Inferring demographic attributes from location check-ins. *In*: *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 295–304.

Appendix A. Effect of Gender, Age and Nationality

To consider whether the demographics of participants has any effect on the sharing decision, we separated the impact of the dimensions from the demographic by averaging the responses of participants for all scenarios and grouping them based on demographics alone. This gives an insight into the degree to which sharing decisions are influenced by age, gender and nationality, as shown in Figure A1. There is clearly a correlation between the sharing decision of participants and their age ($p < 0.0001$), with younger participants more likely to share. The decision to share is also strongly influenced by gender ($p < 0.00001$), with male participants being much more likely to share. Finally, the nationality of participants also had a significant impact ($p < 0.00001$) with participants from Asia and Southern America being more prone to sharing, but European and North American participants showing no strong trend either way.

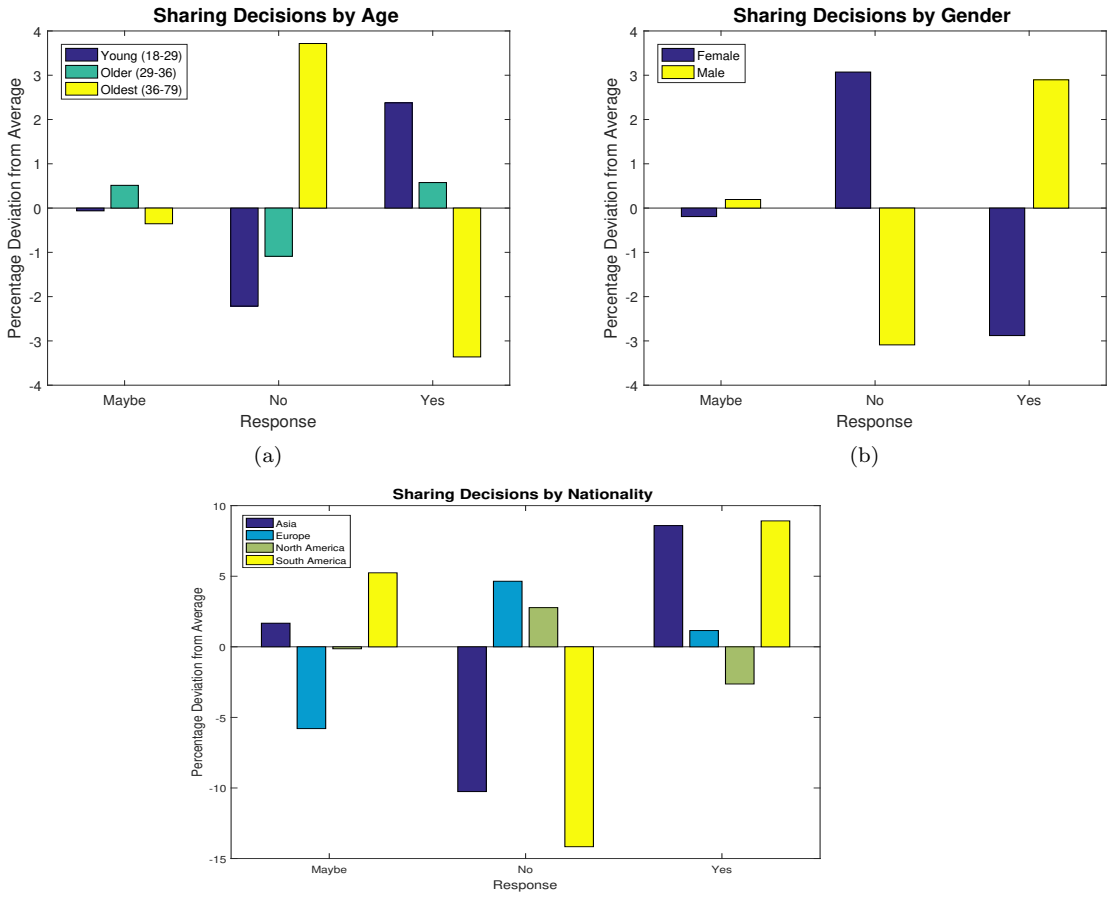


Figure A1.: Effect of gender, age and nationality on sharing decisions. In the case of nationality, Australian and African groups were excluded due to the small sample size (fewer than 30 responses total each).

A.1. *Post-study questionnaire*

A post-study questionnaire, based on the privacy scales proposed in (Rader 2014) was used to gauge how participants value their online privacy. Participants were presented with six statements with a 5-point Likert scale (strongly disagree (1)/strongly agree (5)). The results are shown in Table A1. The first two statements were negatively framed and the rest were positively framed to avoid bias. The average score for each statement was recorded. The overall average score for all statements was 3.845 with a Cronbach's alpha of 0.703, indicating a high degree of consistency in the results.

Table A1.: Post-study questionnaire on the value of personal privacy online.

Statement	**Average**	**SD**
I am not concerned that companies are collecting too much personal information about me.	2.456	1.2
It usually does not bother me when companies ask me for personal information.	2.641	1.17
When people give personal information to a company for some reason, the company should never use the information for any other reason.	4.357	0.92
I have limited the personal information that I post to my social networks' accounts.	3.869	1
I don't post to my social networks' accounts about certain topics because I worry who has access.	3.807	1.14
If I think that information (including location) I posted to my social networks' accounts really looks too private, I might delete it.	4.134	1.04